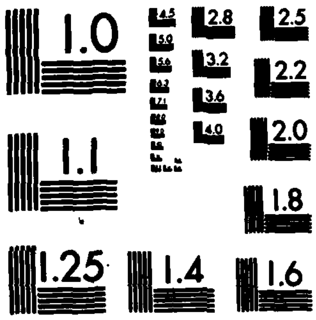
AD-A142 084 THE EFFECT OF INFORMATION DISPLAY FORMAT ON MULTIPLE-CUE JUDGMENT(U) RICE UNIV HOUSTON TX DEPT OF PSYCHOLOGY S P KERKAR ET AL. JUN 84 TR-84-2 1/1

UNCLASSIFIED N00014-82-C-0001 F/G 5/10 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

(12)

RICE UNIVERSITY

Houston, Texas 77001

AD-A142 884

The Effect of Information Display Format on Multiple-Cue Judgment

Shanta P. Kerkar and William C. Howell

Rice University

Technical Report #84-2

June 1984

Department of Psychology

Research Report Series

The Effect of Information Display Format on Multiple-Cue Judgment

Shanta P. Kerkar and William C. Howell

Rice University

Technical Report #84-2

June 1984

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER 84-2	2. GOVT ACCESSION NO. AD.A142 884	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) The Effect of Information Display Format on Multiple-Cue Judgment		5. TYPE OF REPORT & PERIOD COVERED
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Shanta P. Kerkar and William C. Howell		8. CONTRACT OR GRANT NUMBER(s) N00014-82-C-0001
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Psychology Rice University Houston, TX 77251		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR197-074
11. CONTROLLING OFFICE NAME AND ADDRESS		12. REPORT DATE
		13. NUMBER OF PAGES 42
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

policy-capturing, display format, judgment, decision-making, serial processing, holistic processing

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

(see abstract)

DD FORM 1 JAN 73 1473

The Effect of Information Display Format on Multiple-Cue Judgment

Shanta P. Kerkar and William C. Howell

Rice University

ABSTRACT

The rapid evolution of computer technology has drawn considerable attention to the manner in which information is presented on CRT's. Much of the "engineering psychology" research to date has involved evaluation of display parameters (formatting, color, etc.) in terms of performance measures (e.g., speed, accuracy, information rates) that presume a clear-cut criterion. Increasingly, however, users are facing "real world" problems in which no unequivocal criterion exists--situations, for example, in which judgments are required on the basis of displayed data. Since empirical evidence on the effects of display features in these "cognitive" tasks is sparse, three studies were conducted to explore various aspects of the relationship. In all three, subjects were required to combine multiple predictive items (teacher attributes, applicant test scores) into overall evaluations (teacher effectiveness; qualification for a defined position) under conditions of either graphic or numerical display. Using the "policy capturing" methodology, in which multiple regression is used to model behavior, a description of individual judgment strategies was obtained. Display format was found to have a direct influence on the importance attached to (the "weighting of") the separate pieces of information (viz., intelligence etc.) in forming an overall evaluation. Moreover, simultaneous presentation of graphic information tended to produce holistic processing in contrast with the serial processing of numerical information. These findings appear to have important implications for the design of computer-based information processing systems.

INTRODUCTION

Considerable attention has been accorded the coding and formatting of displayed information, particularly in the context of human/computer interface design. Typically, however, the assessment has involved some clearly defined performance criterion viz., speed or accuracy. Thus formats that result in fewer errors or quicker responses are deemed superior to others (see, e.g., Baker & Goldstein, 1966; Cincchinelli & Lantz, 1978; Coffey, 1961; Grace, 1966; Hammond, 1971; Hitt, Schutz, Christner, Ray, & Coffey, 1961; Klemner & Frick, 1953; Tullis, 1981; Wright, 1968). However, there are many applied contexts in which performance is not so easily indexed, notably those involving judgment and/or decision. This is especially true of decisions made under uncertainty since decision outcomes can rarely be used to gauge the quality of the decision making process (Einhorn, 1980; Einhorn & Hogarth, 1981; Lichtenstein, Fishhoff, & Phillips, 1977). In such situations it becomes more meaningful to focus on the _way_ decisions are made (the process itself) rather than what they _are_ (the decision product). Put another way, other aspects of decision performance (e.g., reliability) become more salient than accuracy per se. But, unfortunately, relatively little is known about the effect of format on the decision process. Before examining the rather sparse evidence on this topic, it might be well to review the most commonly used paradigm for studying judgment and decision behavior in the absence of a clear external criterion: the _policy capturing_ paradigm.

The basic approach in policy capturing involves the application of regression analysis to actual judgments in an effort to infer how the individual weights and combines items of predictive information in forming those judgments.[1] Suppose, for example, the judgment of interest was a

commander's evaluation of enemy threat based on surveillance reports, monitored communications, political analyses, and other intelligence information. To determine how much weight he attaches to each predictive item ("cue"), we might ask him to make a series of threat assessments for hypothetical intelligence reports comprised of combinations of the relevant cue values. By regressing the obtained threat judgments on the cue values in a multiple regression analysis, we would describe his weighting "policy". That is, the resulting regression weights would reflect how much importance he tended to attach to the individual cues; the heavier the weight, the greater the importance of that particular cue in determining his judgments. A summary of measures used to describe various aspects of the so captured "policy" is given in Table 1. Of course, the policy capturing approach can also be applied to naturally occurring judgments, although in doing so one must contend with a number of analytic problems (see, e.g., Dawes, 1971; also cf. Ebbesen & Konečni, 1980; Phelps & Shanteau, 1978).

Table 1 about here

The policy capturing paradigm, then, provides a convenient way of describing different aspects of judgment/decision performance in the absence of any accuracy measure. However, despite the increasing presence of computer elements in decision systems, relatively little is known about the effects of display format on individual judgment policies. An early study by Knox and Hoffman (1962) examined the effect of profile format on judgments of intelligence and sociability. The cue values were displayed graphically either as T-scores (with a mean of 50 and standard deviation of 10) or as percentile

scores. It was found that subjects responded "... not only to the underlying meaning of the scores, but to the position of the points on the profile in some absolute sense" (p.19). Extreme cue values produced more variable judgments when expressed as percentile scores than as T-scores (which tended to appear "squeezed in"). Percentile scores resulted in more reliable judgments and higher values of R^2s. However, the regression weights did not differ as a function of format.

In another format comparison, Anderson (1977) found that judgments of teacher quality showed lower linear consistency for verbal than for numerical cue profiles, but that the pattern of weights for the two formats was similar.

One methodological point related to both of these studies concerns the use of standardized regression weights to compare cue weighting under different formats. By definition, standardized weights measure the amount of change in the criterion (Y) in standard deviation (SD) units as a function of one SD unit change in the predictor (Xi). Although independent of scale, expressing change in the criterion in terms of SD units (as standardized weights do) may not always reflect actual changes in cue weighting. Suppose, for example, that a <u>raw score</u> regression weight (bi) increases as a function of some experimental manipulation. Such an increase would produce a direct increase in the SD of the criterion (SDy) as well. But if we then express that weight in standardized form (Bi), the increase in the raw score weight is offset by the increase in SDy, and the real change in cue weighting may not be apparent (see Equation 1).

$$Bi = bi\ (SDx\ /\ SDy) \qquad \text{Equation 1}$$

Lane, Murphy, and Marques (1982) make a convincing argument that the raw score regression weight adjusted by the SD of the <u>cues</u> (SDx) may be a more

appropriate alternative when one wishes to express cue weight independent of the scale of cue measurement but without the problem of a concurrent SDy effect.

Although Anderson (1977) did not report the SD of judgments under the two formats, Knox and Hoffman (1962) found judgments to be more variable in response to percentile scores than to T-scores. It is therefore not immediately obvious whether the failure to find differences in standardized weights in these studies indeed implies that there were no differences in the perceived importance of cues, or whether it was an artifact of the above standardization problem.

Studies of policy capturing thus have offered inconclusive evidence on whether display format affects subjects' cue weighting. Moreover, the comparison formats used were either not readily interpretable (e.g., Knox & Hoffman, 1962) or involved verbal and numerical displays (e.g., Anderson, 1977). The primary goal of the present research, then, was to determine the effect of relatively common display formats on cue weighting and other aspects of the subject's policy when external optimization criteria were not available. The present studies simply compared the effect of two display formats, numerical and graphic representation of cues, on subjects' overall judgments and decisions based on those cues. These formats were chosen because they represent two broad classes of structured displays; moreover, subjects' familiarity with them makes them easily understood. It might be noted that additional task variables were manipulated in conjunction with format in each specific experiment, and that the task scenario was varied between experiments. Scenarios that had "face validity" and that possessed features relevant to the requirements of the experimental paradigm under consideration were chosen.

Further discussion of these methodological features is reserved for the detailed account of each experiment.

EXPERIMENT 1

Given that information regarding display format effects in a policy capturing paradigm is sparse, this experiment was simply an attempt to determine whether a gross format difference (numerical vs. graphic display) would affect either judgments or choices based upon identical input data. Subjects were required to process multidimensional stimuli that were displayed numerically and graphically, and their subsequent responses under both formats were compared.

Subjects performed two types of decision tasks--judgment and choice. A number of investigators have suggested that the type of response required--judgment and choice--influences how people process information (e.g., Einhorn & Hogarth, 1981; Hammond, McClelland, & Mumpower, 1980; Payne, 1982), and thereby produces substantially different kinds of decision behavior. In a judgment task the subject is typically required to assign values to individual alternatives as an expression of psychological worth (e.g., as a rating on a scale or as a representative sum of money he/she would pay for an alternative) whereas in choice the task is to select one or more preferred item(s) from a set of alternatives. For example, evaluating the overall quality of a make and model of an automobile on the basis of information such as size, m.p.g., cost constitutes judgment, whereas selecting a car for purchase from among those available constitutes choice. Both judgment and choice are clearly interdependent in that choosing from a set of alternatives may well entail judging them with respect to several dimensions; nevertheless, making a choice involves explicit consideration of utilities (Edwards & Tversky, 1967), a

dimension that is not necessarily involved in judgment. In view of these considerations, both types of response were examined for possible display effects.

METHOD

<u>Task</u>. The basic tasks required subjects either to rate the suitability of applicants for the job of secretary or to decide whether they should be hired (tasks that most subjects find both meaningful and realistic). More specifically, subjects were presented with profiles of information about hypothetical applicants which were comprised of four dimensions: intelligence, motivation, skill, and experience. Each profile was represented in one of two ways: as a set of numerical scores (numerical format) or as a set of bar graphs (graphic format).

All subjects performed the rating task and the choice task under both display formats. The rating task simply required them to review one applicant profile at a time and rate it on a suitability scale that ranged from 1 (extremely low) to 10 (extremely high). Each subject rated 100 profiles under each format condition. In the choice task, two applicant profiles were presented together and subjects had to indicate which one of the two applicants they considered more suitable for the job. Fifty applicant profile-pairs were presented for choice under each format condition.

<u>Stimuli</u>. Four sets of 100 applicant profiles (designated as p, q, r, and s) were produced by a multivariate normal generator such that values on the four cues were not intercorrelated. A multivariate array of deviates in the range of 0 to 1 was produced. The deviates were further transformed such that the actual values that defined the four cues (intelligence, motivation, skill, and experience) were sampled from populations with means (and SDs) of 25 (15),

5 (2), 10 (3), and 3 (.5) subject to the constraint that the cue values ranged between 1-50, 1-10, 1-25, and 1-5 respectively.

Sets p and q were used in the rating task and sets r and s in the choice task. The profiles in each set were printed on unlined, continuous paper; only one profile with an applicant number appeared on each page in the rating task (sets p and q), but two profiles designated as applicant A and applicant B were presented on the same page in the choice task. Each set of profiles was represented numerically and graphically in separate booklets.

An illustration of the two formats for the rating task is shown in Figure 1. It might be noted that the four cue values were represented as raw scores in the numerical display condition and as standardized scores in the graphic condition. That is, the length of the bar for each graphic cue indicated its

---

Figure 1 about here

---

value on scales adjusted to have comparable physical ranges (see asterisks indicating the upper limit on each scale). This apparent confounding was introduced in an effort to equate scales in terms of their ability to convey the cue values properly. It is often the case that real world situations require processing of numerical cues that are not equated on scale (e.g., GPA and GRE scores in evaluating prospective graduate students) and presenting them in standardized form would have thus destroyed a realistic feature of the task; on the other hand, presenting the raw values in graphic form on scales with radically different ranges would have been confusing from a perceptual standpoint. Of course, the question of which definition of equivalence

(literal or perceptual) is more appropriate is actually an empirical one, and one that was addressed in Experiment 2 (below).

Design. A simple 2 x 2 factorial design was used with display format (numerical vs. graphic) and type of decision task (judgment vs. choice) as within-subjects variables. Both rating and choice tasks under a particular format were performed in a block such that half of the subjects were presented with a block of numerical profiles followed by a graphic block, and the other half performed them in reverse order. The order in which the rating and choice tasks were performed within each of these blocks was also counterbalanced. Half of the subjects were presented with two particular sets of profiles in numerical form (sets p and r) and the other two sets in graphic form (sets q and s) whereas the reverse assignment was used for the remaining subjects. These counterbalancing measures required eight subjects for each replication of the design.

Subjects. Forty-eight subjects were recruited from undergraduate psychology courses at Rice University. They either received $4.00 or course credit in exchange for their participation. An equal number of subjects was assigned randomly to one of the eight conditions.

Procedure. Up to six subjects participated in each experimental session which lasted about an hour. Subjects were given detailed procedural instructions with special attention to the characteristics of the cues and the way they were represented under the two display formats. They then performed the rating and choice task under each format in a sequence determined by the condition to which they were assigned.

Subjects performed the rating task paced by a "beeper" tone that sounded at 8-second intervals; they were allowed 16 seconds for each pair of profiles

in the choice task. Both the rating and choice responses were written on separate response sheets.

RESULTS AND DISCUSSION

The effects of format were evaluated separately for the rating and choice tasks and will be described in turn in the following sections.

Rating Task. Of the 100 profiles rated under each format, 10 at either end were used as buffer profiles, and responses to these were not analyzed. The buffer profiles at the beginning were included to familiarize subjects with the task and to allow them to develop a consistent rating strategy; those at the end were included to reduce any effects of inattentiveness that might occur toward the end of a session (Lane et al., 1982). For every subject, a separate policy equation was obtained for the numerical and graphic displays by regressing each type of judgment on the four cues.

The raw score regression weights obtained for the four cues under a particular display condition serve as an index of the subjects' weighting of those cues for that display. Thus, one way of determining the effect of format on judgment is simply to compare these regression weights using a repeated measures ANOVA. However, since the cues were presented on different scales, it was considered essential to make the raw score weights comparable by multiplying them by the SD of each cue. These adjusted weights represent the magnitude of change in the criterion produced by a change of one SD unit in the predictor (cue) (Lane et al., 1982). The mean adjusted weights are shown in Table 2.

Table 2 about here

The main effect of format was not significant, $\underline{F}(1, 47) < 1$ suggesting that the average weights for all cues combined were comparable for the two formats. Obviously this is less meaningful than the cue x format interaction (which compares weighting policies for the formats); this interaction was highly significant, $\underline{F}(3, 141) = 8.10$, $\underline{p} < .0001$. The main effect of cues was also significant, $\underline{F}(3, 141) = 90.84$, $\underline{p} < .0001$. Clearly, therefore, subjects weighted the four cues differently--intelligence received the highest weight followed by motivation, skill, and experience. And although the average weight for the cues did not differ across formats, the specific weights attached to each cue were more uniform in the graphic than in the numerical display. To explore the nature of this interaction further, individual t-tests were conducted to test for differences in the weights for each of the four cues. The weight for <u>intelligence</u> was reliably smaller for the graphic than for the numerical display, $\underline{t}(47) = 2.29$, $p < .05$; those for <u>motivation</u> and <u>experience</u> were reliably larger, $\underline{t}(47) = 2.60$, $\underline{p} < .02$ and 4.90 $\underline{p} < .01$ respectively. The slight increase in the weight for <u>skill</u> was not significant, $\underline{t}(47) < 1$. What is particularly noteworthy about the observed changes is that the fourth cue, experience, which had virtually no impact on judgment under the numerical format did receive some weight when displayed graphically. Coupled with this increase, the decrease in the highest weight (for intelligence) produced the more even distribution of weights under the graphic display format.

The suggestion that a graphic format encourages subjects to consider and weight <u>all</u> cues whereas numerical presentation restricts their attention to a subset of cues must, of course, be tempered by the fact that format and scale representation were confounded in this study (see METHOD). It will be recalled

that the numerical cues were presented as raw scores, whereas the graphic cues were presented in standardized form on physically identical scales. Such rescaling of cues directly affected their variability. Consequently cues with lower variability (viz., motivation, skill, and experience) may have appeared to be more "scattered" in the graphic format thereby inflating their cue weights relative to the numerical format. This possibility, of course, represents an alternative explanation for the results--particularly with respect to the finding that experience, which had the lowest variance, was weighted substantially more heavily under the graphic than the numerical display--which was addressed directly in Experiment 2.

Besides regression weights (which index subjects' cue utilization), another useful descriptive measure is the linear consistency of individual policies viz., the squared multiple correlation or R^2s obtained from regressing judgments on the four cues. The overall difference between R^2s obtained from the numerical and graphic policies (0.64 vs. 0.69) was not significant, $F(1, 47) = 2.69$, $p > .10$. However it should be noted that the variances in R^2s can be partitioned as follows:

$$R^2s = SS\hat{y} \,/\, SSy \qquad \text{Equation 2}$$

but,

$$SSy = SS\hat{y} + SSe \qquad \text{Equation 3}$$

so

$$R^2s = SS\hat{y} \,/\, (SS\hat{y} + SSe) \qquad \text{Equation 4}$$

Thus a comparison of the two formats in terms of the $SS\hat{y}$ and SSe components was deemed more meaningful than the overall R^2s index. Since SS measures tend to have skewed distributions, a square-root transformation was applied to both $SS\hat{y}$ and SSe for purposes of analysis. Resulting t-tests showed

that mean $\sqrt{SS\hat{y}}$ values were not significantly different under the two display conditions: 12.19 (numerical) vs. 11.61 (graphic), $t(47) = 1.07$, $p > .10$. On the other hand, $\sqrt{SSe}$ differences were highly significant: 8.90 (numerical) vs. 7.61 (graphic), $t(47) = 3.59$, $p < .001$. What this finding suggests is that the graphic format produced considerably more <u>precision</u> in judgment than did the numerical format, a conclusion that is reinforced by the fact that variability in raw criterion judgments was also significantly lower for the graphic display: SD = 1.58 vs. 1.74, $t(47) = 3.63$, $p < .001$.

The analyses discussed so far assume that subjects' policies could be described adequately in terms of a linear model. Since occasional instances of nonlinearity have been reported (e.g., Einhorn, 1970; Einhorn, 1971; Wiggins & Hoffman, 1968), a quadratic and a configural model were also applied to subjects' judgments. Both models included as predictors the four cues (Xi) and the coded format vector. In addition, the quadratic model included the four squared values of cues (Xi^2) and their interactions with the coded vector and the configural model included 11 cross-products of cues (XiXj) and their interactions with the coded vector. Results of these analyses indicated some nonlinearity for a few subjects. However, even those who showed significant nonlinearity could be described adequately in terms of a linear model--the linear model alone accounted for 91.90% and 90.80% of the total variance accounted for by the quadratic and configural models respectively.[2]

<u>Choice Task</u>. The primary question of interest here was whether choice performance differed significantly with format. Since there was no <u>external</u> criterion available to define choice accuracy, the subjects' own numerical and graphic rating policies were used as criteria. That is, "policy captured" weights were applied to the cue values for each pair of choice profiles to

determine which profile should be chosen if the individual were consistent with his/her own policy. These predicted choices were then compared to actual choices under the two formats to obtain "accuracy" measures. Since there were two policies (numerical and graphic) for each set of values, it was also possible to compare decision "accuracy" for consistent criteria (e.g., actual numerical choices evaluated with reference to a numerical policy) with those for inconsistent criteria (e.g., actual graphic choices evaluated against a numerical policy). These "accuracy" scores were analyzed in a 2 x 2 ANOVA design with format and consistency of rating policy as the two within-subjects variables. The mean accuracy scores are reported in the first two rows of Table 3.

Table 3 about here

Neither the effect of format nor the interaction between format x consistency of policy was significant, $F(1, 47) = .58$ and 1.60, $p = .45$ and $.21$ respectively. This suggests that despite the differences in subjects' rating policies under the two formats, they predicted choices with similar levels of accuracy. There was, however, a significant effect of consistency, $F(1, 47) = 9.43$, $p < .01$. Although the absolute differences were extremely small, a consistent policy predicted slightly better than an inconsistent one. This implies that subjects' rating and choice behavior were more similar when information was displayed in identical than in different formats. Thus while numerical and graphic cues were processed differently, the same display mode induced similar kinds of processing for both rating and choice tasks.

Since some have argued that, from a practical standpoint, judgment is predicted as well applying unit weights to the cues as it is with "policy captured" or derived "importance" weights (Dawes & Corrigan, 1974; Dawes, 1979; Einhorn & Hogarth, 1975), it is of interest to compare the efficacy of the two models for the present data. Hence predictions from a unit-weighted model were obtained separately for the numerical and graphic profiles and compared to subjects' actual choices under those formats. The resulting accuracy scores are reported in the third row of Table 3. For both graphic and numerical formats, the consistent rating policy predicted better than a unit-weighted model, $t(47) = 5.06$ and 2.66, $p < .001$ and $< .05$ in each case. However, the inconsistent policy predicted better than the unit-weighted model only for the graphic format, $t(47) = 2.35$, $p < .05$; for the numerical format, the difference was not reliable, $t(47) = 1.95$, $p > .05$. The finding that the consistent policy for both formats fared reliably better than the inconsistent one corroborates the conclusion that subjects processed stimulus profiles similarly (regardless of task) under a particular display. It also implies that the regression policy did capture something important about the subject's behavior under a particular display format.

In summary, the major conclusion to be drawn from this study is that display format does induce differences in the way people handle predictive data, although as one might suspect, the processes involved are not necessarily simple.

EXPERIMENT 2

The findings of Experiment 1 suggested a difference in pattern of cue weighting for numerical and graphic formats. More specifically, there was a tendency for the graphic format to produce a more even weighting of cues than

the numerical format. Whether this effect was the result of the display format per se or the confounded difference in representation of scale values, however, remained unclear (see earlier discussion). Of course, scale features are themselves an aspect of display formatting, although it was not the aspect to which Experiment 1 was primarily addressed.

Therefore, Experiment 2 sought to remove the confounding of scale with display format effects. The design was similar to that of Experiment 1 except that it was limited to the rating task, and all cues were represented on comparable scales. The main purpose of this experiment, then, was to evaluate the effect of format on the judgment of otherwise strictly equivalent information.

METHOD

Materials and Design. Subjects performed a rating task similar to the one described in Experiment 1. They reviewed profiles of hypothetical applicants for the job of secretary and rated them on a suitability scale that ranged from 1 (extremely low) to 10 (extremely high). Each profile contained information about the applicant's intelligence, motivation, social skill, and typing ability.

Two sets of 100 profiles (p, q) were generated in a manner identical to that in Experiment 1 except that values on all 4 dimensions--intelligence, motivation, skill, and typing ability--were sampled from populations with a mean of 25 and SD of 15. The scores were generated randomly subject to the constraint that they ranged between 1-50. Both profile sets were represented numerically and graphically.

The format in which the information was displayed was varied within subjects: under the numerical format the cue values were presented as

numerical scores, whereas under the graphic format they were presented as horizontal bar graphs (refer to Figure 1). The presentation of the cues under the numerical and graphic format was the same as Experiment 1, except that all cues were represented on comparable scales and thus no standardization was necessary for the graphic display. The sequence in which the numerical or graphic information was displayed was counterbalanced such that half of the subjects rated numerical followed by graphic profiles and the other half rated them in the reverse order. The specific sets of profiles (p and q) used in the numerical and graphic conditions were rotated so that they were represented in both formats. Such counterbalancing resulted in four different conditions.

Subjects. Twenty Rice University students served in the experiment for course credit toward undergraduate psychology courses or for pay. They participated in the experimental sessions individually or in groups of 3-6. Subjects were randomly assigned to the four conditions under the constraint that an equal number of subjects appeared in each condition.

Procedure. After initial instructions regarding the task, subjects received individual booklets in which the profiles to be rated were printed. The sequence in which the numerical and graphic profiles were presented and also the specific set of profiles reviewed was determined by the condition to which the subject was assigned. Subjects reviewed and rated each of the 100 profiles in the numerical and graphic format at the rate of 8 seconds per profile. A 5 minute rest interval was interposed between the rating of the two sets.

RESULTS AND DISCUSSION

As in Experiment 1, subjects rated 100 profiles under each format and judgments for 10 buffer profiles at either end were not analyzed. Thus every

subject's policy equation for the two formats was based on judgments to the remaining 80 profiles. The mean raw score regression weights obtained through policy capturing are shown in Table 4. An ANOVA applied to these weights showed a marginal effect of format, $\underline{F}(1, 19) = 3.68$, $\underline{p} = .07$; a significant effect of cue, $\underline{F}(3, 57) = 4.33$, $\underline{p} = .008$; and a significant cue x format interaction, $\underline{F}(3, 57) = 2.88$, $\underline{p} = .04$.

Table 4 about here

These results replicate the primary finding of Experiment 1--format again produced a differential weighting of cues. Thus subjects do indeed process the same cues differently under the numerical and graphic display and the obtained interaction does not seem to be dependent specifically on the scaling features peculiar to those in Experiment 1. In order to describe this interaction precisely, individual t-values were obtained by comparing the mean weights for each cue under the two formats. Only one was significant: the weight for motivation displayed graphically was reliably larger than that displayed numerically, $\underline{t}(19) = 2.53$, $\underline{p} < .05$; those for intelligence , skill, and typing ability all failed to achieve significance, $\underline{t}(19) = 1.05$, 1.87, and 1.46, $\underline{p} > .05$ in each case. It will be noted that there were a number of shifts in mean weigting of cues from Experiment 1 (see Tables 2 and 4). Numerical cues received a higher average weight in this experiment, a fact that could be attributed to an ease of processing of cues due to equated scale units; this point is obviously not applicable to the graphic format since all cues were presented on comparable scales in both experiments.[3] Consequently, these shifts in mean weights obscure any tendency for graphic cues to produce

more even weighting than numeric ones. We can thus merely note that in this experiment, the weighting of individual cues differed with format, but in no simply described pattern.

Turning to the consistency of judgments (R^2s), numerical policies were less consistent than graphic ones (0.54 vs. 0.67); this difference was reliable, $\underline{F}(1, 19) = 8.13$, $\underline{p} < .01$. R^2s was broken down into its two components, $SS\hat{y}$ and SSe, and a separate comparison of $\sqrt{SS\hat{y}}$ and $\sqrt{SSe}$ was made for the two formats. The mean $\sqrt{SS\hat{y}}$ for the numerical and graphic judgments were 12.12 and 13.72, with a t-test on this difference showing $\underline{t}(19) = 1.86$, $\underline{p} < .10$; $\sqrt{SSe}$ for numerical judgments was larger then that for graphic (11.21 vs. 9.47), $\underline{t}(19) = 2.57$, $\underline{p} < .02$. It thus appears that the lower consistency of numerical judgments resulted largely from greater error in those judgments than in the graphic ones. The finding that the numerical format produced lesser precision than the graphic one parallels that of Experiment 1; however, the lower precision did not affect the SDs of the judgments (1.90 vs. 1.89 for numerical and graphic formats, $\underline{t}(19) < 1$). A summary of R^2s and its component measures for the two Experiments is provided in Table 5.

Table 5 about here

Summarizing the first two experiments, it appears that format does influence the manner in which people weight cues, but the nature of this influence is not simply described. It may, in fact, be quite idiosyncratic. Nonetheless, one generalization does emerge: judgment is less consistent under the numerical format, and this is attributable chiefly to the lower precision of numerical judgments relative to graphic ones.

EXPERIMENT 3

This experiment was simply an attempt to elucidate further the underlying nature of the differences produced by the numerical and graphic formats. Research on the effect of structural properties of stimuli on perceptual tasks has suggested that some stimulus dimensions are perceived holistically (integral dimensions), while others are perceived individually (separable dimensions) (Garner, 1974). For example, the height and width of a rectangle are combined holistically to produce perception of rectangular area (Felfoldy, 1974; Garner & Felfoldy, 1971; Lockhead, 1979). Later investigations have generalized this result to include decision tasks, more particularly in the multiple cue probability learning paradigm in which subjects acquire knowledge regarding cue-criterion relations (Goldsmith & Schvaneveldt, 1982; Wickens & Scott, 1983).

It appears, then, that the graphic format might encourage a <u>holistic</u> perception of cues presented together; the numerical format, however, might produce serial processing. It was postulated that the simultaneous presentation of cues in the first two experiments may have favored a holistic processing of graphic cues. If, then, cues were presented sequentially rather than simultaneously, the holistic perception of graphic cues would be largely eliminated and, as a consequence, so would the difference between the numerical and graphic formats.

The present experiment, therefore, involved a sequential presentation of cues under numerical and graphic formats. One additional manipulation involved the number of cues presented to subjects. Previous investigators (e.g., Einhorn, 1971) have found that increasing the "information load" increases the difficulty of integrating cues and thus is detrimental to performance. In a

sequential presentation of cues, subjects are obliged to rely heavily on memory in making their judgments or choices, thus exacerbating the difficulty. By varying the number of cues, therefore, the interest was to provide an adequate range of task difficulty for the appearence of any potential format effects.

METHOD

Materials and Design. As in Experiments 1 and 2, subjects rated multidimensional stimuli on a global dimension. However, the present task involved teaching effectiveness judgments instead of the personnel selection/rating tasks used previously. The primary reason for this change was to explore the generality of display effects in another realistic judgment context, while preserving the formal properties of the task. The stimuli consisted of profiles of hypothetical college instructors whose performance was described with respect to either four or six cues, and values of the cues were displayed either numerically or graphically. The design, then, involved the factorial combination of two variables--number of cues (four or six) which was manipulated between-subjects and display format (numerical vs. graphic) which was manipulated within-subjects.

A set of 200 profiles was generated in a manner identical to that described in Experiment 1 except that (1) six cue values were generated per profile, and (2) all cues were sampled from populations with a mean of 30, a SD of 10, and a range of 1-60. The six cues describing the profiles were designated as information imparted in course, arousal of interest, presentation style, knowledge of the field, rapport with students, and clarity of course requirements. Subjects in the four-cue condition were presented with a subset of these six cues; however, the exact subset of cues was sampled independently for each subject. The order in which information

under the two formats was displayed was counterbalanced in both the four- and six-cue conditions. Thus half of the subjects rated numerical profiles followed by graphic profiles and the reverse was true for the remaining subjects.

The numerical and graphic presentation of cues was similar to Experiments 1 and 2. As in Experiment 2, all cues had identical scale units so that no transformation of the cues was necessary for graphic presentation. For every subject 100 profiles were chosen randomly for graphic presentation and the remaining 100 profiles were presented numerically. The order of presentation of the cues (for the four- and six-cue conditions) and the selection of the specific four-cue subsets (for the four-cue condition) were also randomized individually. Any given subject, however, reviewed the same cues in a specific order under both types of display format.

Subjects. The experiment was conducted in individual sessions that lasted for about an hour. Twenty subjects, enrolled in undergraduate psychology courses at Rice University, participated in the experiment in exchange for course credit or pay.

Procedure. The subject was seated in a cubicle before the screen of a TRS-80 (Model II) microcomputer. After initial instructions regarding the task and procedure, the profiles of instructors were displayed on the screen of the computer, one profile at a time. Each profile was presented in the following manner. First the words "Instructor #" appeared on the screen along with the number of the profile being rated. This message served primarily as a preparatory signal for subjects to attend to the incoming information and also to distinguish one profile from another. Then the cues were presented successively, at a 2 second rate, each with its label and value. After all

four or six cues were presented, the instructions "Rate the instructor on a scale of 1 to 10" were displayed on an otherwise blank screen, and the subject proceeded to write down his/her rating on a separate response form. After the response had been recorded, the experimenter depressed a programmed key on the computer keyboard to initiate the next profile. Thus, although the time of presentation of cues was controlled, subjects' responses were essentially self-paced. A brief rest period intervened between the ratings of two sets of profiles.

RESULTS AND DISCUSSION

The principal issue addressed in this experiment was the effect of sequential presentation of cues on subjects' judgments under numerical and graphic display formats. The expectation was that the sequential procedure would eliminate the display effect if, in fact, the primary causative factor was holistic processing.

Of the 100 profiles rated under each format, ratings of 10 buffer profiles at either end were not analyzed. Both numerical and graphic policies for each subject were then obtained by regressing the 80 judgments on four or six cues. The raw score regression weights from the numerical and graphic policies were then analyzed in two ways: for order effects and for specific cue effects.

The first type of analysis pertained to the weights attached to the sequential position of successively presented cues. Note that order of presentation of specific cues (e.g., information imparted in course or presentation style) was randomized individually so that the effect of cue in this analysis does not pertain to a particular cue across subjects. Separate

ANOVAs were applied to the four- and six-cue data which are presented in Table 6.

Table 6 about here

Looking first at the four-cue ANOVA, the means for the numerical and graphic formats were .64 and .52 respectively, a difference that was significant, $F(1, 9) = 11.01$, $p < .01$. Thus, cues tended to be weighted more heavily <u>on average</u> under the numerical than the graphic display. However, as predicted, the cue x format interaction did not approach significance, $F(3, 27) < 1$. Given that this interaction was highly significant under a simultaneous presentation of cues in both Experiments 1 and 2, the failure to find it in the present data lends indirect support to the hypothesis that graphic display encourages holistic processing. But obviously, this conclusion must be considered tentative due to the inherent danger in accepting the null hypothesis. The effect of cue was only marginally significant, $F(3, 27) = 2.55$, $p = .08$.

The six-cue ANOVA also failed to reveal a significant cue x format interaction, $F(5, 45) = 1.12$, $p = .37$, thereby supporting the claim that a sequential presentation eliminated the holistic processing of graphic cues. However, the main effect of format found in the four-cue condition was absent here, $F(1, 9) < 1$. Exactly why this should occur is not clear. There was also no suggestion of the presence of cue effects, $F(5, 45) = 1.23$, $p = .31$.

The analyses discussed so far determined whether the processing of cues was affected by the format in which they were presented and their temporal ordering. The second analytic approach was based on cues irrespective of order, the purpose being to establish whether the particular <u>cues</u> were

weighted differently under the two formats. This analysis was possible only under the six-cue condition since the four-cue condition did not provide all subjects with the same subsets of cues. Since, the effect of format or the cue x format interaction were not significant, $F(1, 9) < 1$ and $F(5, 45) = 1.07$, $p = .39$, the regression weights were collapsed across format and these means are presented in Table 7. As is apparent from Table 7, there was clearly an effect of cue, $F(5, 45) = 3.77$, $p = .006$. As one would expect, some cues were weighted more heavily than others.

Table 7 about here

In sum, there was no evidence for a differential weighting of cues presented sequentially under the two formats--the cue x format interaction found consistently in the first two experiments was eliminated in this one. This supports our hypothesis of holistic processing of graphic cues. However, there were some processing differences as a function of format; the numerical format produced larger overall cue weights than the graphic format in the four-cue condition.

The consistency or R^2s obtained from subjects' policies was compared in a 2 x 2 ANOVA design, with number of cues (four vs. six cues) as a between-subjects variable and format (numerical vs. graphic format) as a within-subjects variable. The mean R^2s for the four- and six cue conditions was .67 and .57 respectively and the decline in consistency as the number of cues increased from four to six was significant, $F(1, 18) = 8.74$, $p = .008$. However, neither the effect of format nor the number of cues x format interaction was significant, $F(1, 18) = 1.90$ and 1.62, $p = .19$ and $.22$.

The finding that an increase in information load affects R^2s is supported by previous studies (e.g., Anderson, 1977; Billings & Marcus, 1983; Einhorn, 1971). However, lowered consistency could result either from a decrease in cue usage due to the greater amount of processing load imposed by additional cues (measured by $SS\hat{y}$) or an increase in random error (measured by SSe). Looking at these components, only $\sqrt{SS\hat{y}}$ differed for the four- and six-cue conditions (11.12 vs. 8.92), $F(1, 18) = 4.33$, $p = .05$; the difference between $\sqrt{SSe}$ was not reliable (7.74 vs. 7.58), $F(1,18) < 1$. Format did not affect $\sqrt{SS\hat{y}}$ or $\sqrt{SSe}$ significantly and the number of cues x format interaction also failed to approach significance for both measures. These findings suggest that subjects who had a larger set of cues to process (six-cue condition) tended to use the information less completely than those who had a smaller set (four-cue condition), consequently lowering the linear consistency of their policies. Whether the sequential presentation of cues imposed an additional memory load and caused a greater decrement in consistency for the former condition relative to a simultaneous presentation is not possible to determine from these data.

GENERAL DISCUSSION

Two common formats for displaying "cue information" were compared with respect to their influence on judgment and choice behavior under different task scenarios. The most important finding was that subjects weighted the same cues differently when displayed numerically than they did when displayed in graphic form. That is, their judgments and choices suggested that they attached consistently more (or less) importance to particular items of information under one format than under the other, and they did so irrespective of the task scenario used (e.g., whether the judgment involved the suitability of job candidates or the evaluation of instructors).

These differences disappeared, however, under conditions of <u>sequential</u> cue presentation (Experiment 3), a situation designed to minimize the holistic processing tendency believed to occur with the graphic format. Thus a necessary condition for the demonstration of format-induced differences is the simultaneous availability of cue values. Presumably, people tended to process numerical information serially in any case, while they may operate in a more holistic (simultaneous processing) mode if multiple graphic inputs are available simultaneously. Additional evidence for the holistic processing hypothesis was obtained in Experiment 1, where the graphic display produced more uniform cue weightings than did the numerical display. However, the Experiment 2 results were equivocal with respect to this tendency, so the question of whether a graphic format encourages the operator to pay more attention to more of the available cues is still an open one. Although holistic processing would seem logically to encourage more complete use of predictive information, it does not follow that the resulting weights must be more <u>uniform</u> than for serial processing.

The present evidence for display-induced differences in cue weighting is contrary to several previous reports (e.g., Anderson, 1977; Goldsmith & Schvaneveldt, 1982; Knox & Hoffman, 1962; Wickens & Scott, 1983). The difference may well reflect an important methodological point regarding the calculation of regression weights. As noted earlier, the conventional method uses <u>standardized</u> weights that may be insensitive to actual changes in cue weighting. Consequently, a more powerful alternative measure--the <u>raw score</u> regression weight--was used in the present analyses. The fact that it revealed significant differences whereas the previous research had not lends credence to the argument that the raw score weight is more sensitive and hence more appropriate than standardized weights for measuring cue importance (Lane et al., 1982).

One other methodological point that has been virtually ignored in previous research concerns R^2s or the consistency of subjects' policies. R^2s refers to the proportion of variance in actual judgments that is accounted for by the variance in predicted judgments (that are based on a weighted combination of cues). Typically, studies have reported either R^2s alone (e.g., Einhorn, 1971; Knox & Hoffman, 1962) or R^2s plus the variance of criterion judgments without explicating their relationship (e.g., Anderson, 1977). As was illustrated by the present research, important information concerning the effect of experimental manipulations can be revealed by examining the components of R^2s--the variance of actual judgments (or, alternatively, SSy), the variance of predicted judgments (or, $SS\hat{y}$), and variance of error (or, SSe). Thus in Experiment 2, for example, the lower consistency of numerical vs. graphic policies was shown to stem largely from the higher magnitude of error (SSe) in those judgments. Since the components have independent meaning, it is obvious that more can be learned about the underlying processes by analyzing $SS\hat{y}$ and SSe separately than by merely reporting overall consistency (R^2s).

In sum, the present research serves to demonstrate that format can affect judgment and choice behavior, although the precise nature of the processing difference was not established. While the results were generally consistent with a holistic-serial processing distinction, they did not prove the point. The fact that the present findings failed to confirm null format results reported elsewhere is attributed to use of insensitive "importance weighting" measures in that research. This, as well as several other methodological refinements for studying judgment performance were developed and illustrated in the three reported experiments.

FOOTNOTES

1. Traditionally the multiple regression model in policy capturing has been used for the primary purposes of identifying the underlying judgment "process" and/or predicting judgment "outcomes". As has been argued elsewhere (see Kerkar, 1983), the usefulness of the paradigm can be enhanced considerably if it is applied in a functional manner. Very simply, within a functional framework the regression model is used to index performance in decision tasks with varying demands: the goal is to relate task features to behavioral consequences without undue emphasis on modeling processes or capturing outcomes. The regression model has been used within such a functional framework in the experiments reported here.

2. There was no evidence that nonlinearity varied systematically with display format. Since a similar pattern of data was observed for Experiments 2 and 3, a discussion of these results is omitted.

3. This observation overrides any differences in cue weighting that might arise from changes in cue <u>labels</u> from Experiment 1 to Experiment 2.

REFERENCE NOTES

Kerkar, S. P. A critical analysis of the uses of multiple regression in the study of human judgment. Technical Report # 83-2. Department of Psychology Research Report Series, Rice University, 1983.

Wickens, C. D., & Scott, B. D. A comparison of verbal and graphical information presentation in a complex information integration decision task. Technical Report No. EPL-83-1/ONR-83-1. Engineering-Psychology Research Laboratory, University of Illinois at Urbana-Champaign, 1983.

REFERENCES

Anderson, B. L. Differences in teachers' judgment policies for varying numbers of verbal and numerical cues. Organizational Behavior and Human Performance, 1977, 19, 68-88.

Baker, J. D., & Goldstein, I. Batch vs. sequential displays: Effects on human problem solving. Human Factors, 1966, 8, 225-235.

Billings, R. S., & Marcus, S. A. Measures of compensatory and noncompensatory models of decision behavior: Process tracing versus policy capturing. Organizational Behavior and Human Performance, 1983, 31, 331-352.

Cincchinelli, L. F., & Lantz, A. E. The elimination of irrelevant information in a multidimensional stimulus classification task. Human Factors, 1978, 20, 78-82.

Coffey, J. L. A comparison of vertical and horizontal arrangements of alphanumerical material--Experiment 1. Human Factors, 1961, 3, 93-98.

Dawes, R. M. A case study of graduate admissions: Application of three principles of human decision making. American Psychologist, 1971, 26, 180-188.

Dawes, R. M. The robust beauty of improper linear models in decision making. American Psychologist, 1979, 34, 571-582.

Dawes, R. M., & Corrigan, B. Linear models in decision making. Psychological Bulletin, 1974, 81, 95-106.

Ebbesen, E. B., & Konečni, V. On the external validity of decision-making research: What do we know about decisions in the real world. In T. S. Wallsten (Ed.), Cognitive Processes in Choice and Decision Behavior. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1980.

Edwards, W., & Tversky, A. (Eds.). Decision Making. Penguin, 1967.

Einhorn, H. J. The use of nonlinear, noncompensatory models in decision making. Psychological Bulletin, 1970, 73, 221-230.

Einhorn, H. J. Use of nonlinear, noncompensatory models as a function of task and amount of information. Organizational Behavior and Human Performance, 1971, 6, 1-27.

Einhorn, H. J. Learning from experience and suboptimal rules in decision making. In T. S. Wallsten (Ed.), Cognitive Processes in Choice and Decision Behavior. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1980.

Einhorn, H. J., & Hogarth, R. M. Unit weighting schemes for decision making. Organizational Behavior and Human Performance, 1975, 13, 171-192.

Einhorn, H. J., & Hogarth, R. M. Behavioral decision theory: Processes of judgment and choice. Annual Review of Psychology, 1981, 32, 53-88.

Felfoldy, G. L. Repetition effects in choice reaction time to multidimensional stimuli. Perception and Psychophysics, 1974, 15, 453-459.

Garner, W. R. The Processing of Information and Structure. New York: John Wiley & Sons, 1974.

Garner, W. R., & Felfoldy, G. L. Integration of stimulus dimensions in various types of information processing. Cognitive Psychology, 1971, 1, 225-241.

Goldsmith, T. E., & Schvaneveldt, R. W. The role of integral displays in decision making. Proceedings of Human Factors in Computer Systems, 1982.

Grace, G. L. Application of empirical methods to computer-based system design. Journal of Applied Psychology, 1966, 50, 442-450.

Hammond, K. R. Computer graphics as an aid to learning. Science, 1971, 172 (3986), 903-907.

Hammond, K. R., McClelland, G. H., & Mumpower, J. Human Judgment and Decision Making: Theories, Methods, and Procedures. New York: Praeger, 1980.

Hitt, W. D., Schutz, H. G., Christner, C. A., Ray, H. W., & Coffey, L. J. Development of design criteria for intelligence display formats. Human Factors, 1961, 3, 86-92.

Klemner, E. T., & Frick, F. C. Assimilation of information from data and matrix patterns. Journal of Experimental Psychology, 1953, 45, 15-19.

Knox, R. E., & Hoffman, P. J. Effects of variation of profile format on intelligence and sociability judgments. Journal of Applied Psychology, 1962, 46, 14-20.

Lane, D. M., Murphy, K. R., & Marques, T. E. Measuring the importance of cues in policy capturing. Organizational Behavior and Human Performance, 1982, 30, 231-240.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. Calibration of probabilities: The state of the art. In H. Jungermann & G. deZeeuw (Eds.), Decision Making and Change in Human Affairs. Dordrecht-Holland: D. Reidel Publishing Co., 1977.

Lockhead, G. R. Holistic vs. analytic process models: A reply. Journal of Experimental Psychology: Human Perception and Performance, 1979, 5, 746-754.

Payne, J. W. Contingent decision behavior. Psychological Bulletin, 1982, 92, 382-402.

Phelps, R. H., & Shanteau, J. Livestock judges: How much information can an expert use? Organizational Behavior and Human Performance, 1978, 21, 209-219.

Tullis, T. S. An evaluation of alphanumeric, graphic, and color information displays. Human Factors, 1981, 23, 541-550.

Wiggins, N., & Hoffman, P. J. Three models of clinical judgment. Journal of Abnormal Psychology, 1968, 73, 70-77.

Wright. P. Using tabulated information. Ergonomics, 1968, 11, 331-343.

TABLE 1

Descriptive measures obtained from policy capturing

bi =	raw score regression weight for cue 'i' obtained from the subject's policy, indicating the importance attached to that cue
R^2s =	linear consistency of the subject's policy
$SS\hat{y}$ =	sum of squared deviations of the subject's predicted responses ($\hat{Y}s$) from their mean
SSe =	sum of squared deviations of the residual in the subject's responses that could not be predicted from a weighted combination of the cues ($Ys - \hat{Y}s$) from their mean
SSy =	sum of squared deviations of the subject's responses (Ys) from their mean or ($SS\hat{y}$ + SSe)

Table 2

Mean raw score regression weights (adjusted) for the four cues under the numerical and graphic formats (Experiment 1).

	Cues			
	Intelligence	Motivation	Skill	Experience
Numerical Format	1.12	.57	.37	.03
Graphic Format	.97	.70	.33	.15

Note: Adjusted regression weights were obtained by multiplying the raw score weights and standard deviations of cue values to equate scale differences among the four cues.

Table 3

Choice Accuracy based on a comparison of subjects' actual choices under the two formats and choices predicted from their regression models and a unit-weighted model.

	Graphic Format	Numerical Format
Consistent rating policy	83.96	83.54
Inconsistent rating policy	80.63	82.86
Unit-weighted model	77.55	80.42

Table 4

Mean raw score weights for the four cues under the numerical and graphic formats (Experiment 2).

	Cues			
	Intelligence	Motivation	Social Skill	Typing Ability
Numerical Format	.82	.73	.46	.78
Graphic Format	.75	.94	.57	.93

Note: Since the scales had equal SDs, no adjustment was necessary for the analyses. The means in the Table are, however, multiplied by the SDs to make them comparable to those of Experiment 1.

Table 5

Summary of R^2s and its component measures from Experiments 1 and 2

	Experiment 1		Experiment 2	
	Numerical	Graphic	Numerical	Graphic
R^2s	.64	.69	.54	.67
SDy	1.74	1.58	1.90	1.89
$\sqrt{SS\hat{y}}$	12.19	11.61	12.12	13.72
$\sqrt{SSe}$	8.90	7.61	11.21	9.47

Table 6

Mean raw score regression weights for the successively presented cues.

Number of Cues	Format	Order of Presentation					
		1	2	3	4	5	6
Four	Numerical	.55	.51	.66	.85	--	--
	Graphic	.38	.45	.57	.67	--	--
	$\bar{X}$	.46	.48	.61	.76	--	--
Six	Numerical	.35	.37	.24	.30	.54	.36
	Graphic	.29	.42	.28	.29	.46	.38
	$\bar{X}$	.32	.40	.26	.30	.50	.37

Note: The means in the Table are adjusted by the SDs of the cues to make the data consistent with those from Experiments 1 and 2.

Table 7

Mean raw score regression weights (collapsed across format) for the specific cues in the six-cue condition.

Information imparted in course	Arousal of interest	Presentation style	Knowledge of field	Rapport with students	Clarity of requirements
.47	.56	.35	.32	.25	.19

Numerical Display

Intelligence	35
Motivation	6
Skill	16
Experience	3

Graphic Display

```
IQ      ***********************************                *
MOTIV   ******************************                     *
SKILL   ********************************                   *
EXP     ******************************                     *
```

FIGURE 1. Illustration of the numerical and graphic formats.

MAY 9 1984

May 1984

OFFICE OF NAVAL RESEARCH

Engineering Psychology Group

TECHNICAL REPORTS DISTRIBUTION LIST

OSD

CAPT Paul R. Chatelier
Office of the Deputy Under Secretary
of Defense
OUSDRE (E&LS)
Pentagon, Room 3D129
Washington, D. C. 20301

Dr. Dennis Leedom
Office of the Deputy Under Secretary
of Defense (C^3I)
Pentagon
Washington, D. C. 20301

Department of the Navy

Engineering Psychology Group
Office of Naval Research
Code 442EP
800 N. Quincy St.
Arlington, VA 22217 (3 cys.)

Aviation & Aerospace Technology
Programs
Code 210
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

CDR. Paul Girard
Code 250
Office of Naval Research
800 North Quincy Street
Arlington, VA 22211

Physiology Program
Office of Naval Research
Code 441NP
800 North Quincy Street
Arlington, VA 22217

Dr. Edward H. Huff
Man-Vehicle Systems Research Division
NASA Ames Research Center
Moffett Field, CA 94035

Deparment of the Navy

Dr. Andrew Rechnitzer
Office of the Chief of Naval
Operations, OP952F
Naval Oceanography Division
Washington, D.C. 20350

Manpower, Personnel & Training
Programs
Code 270
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Mathematics Group
Code 411-MA
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Statistics and Probability Group
Code 411-S&P
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Information Sciences Division
Code 433
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

CDR Kent S. Hull
Helicopter/VTOL Human Factors Office
NASA-Ames Research Center MS 239-21
Moffett Field, CA 94035

Dr. Carl E. Englund
Naval Health Research Center
Environmental Physiology
P.O. Box 85122
San Diego, CA 92138

Department of the Navy

Special Assistant for Marine Corps
Matters
Code 100M
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Mr. R. Lawson
ONR Detachment
1030 East Green Street
Pasadena, CA 91106

CDR James Offutt, Officer-in-Charge
ONR Detachment
1030 East Green Street
Pasadena, CA 91106

Director
Naval Research Laboratory
Technical Information Division
Code 2627
Washington, D.C. 20375

Dr. Michael Melich
Communications Sciences Division
Code 7500
Naval Research Laboratory
Washington, D. C. 20375

Dr. J. S. Lawson
Naval Electronic Systems Command
NELEX-06T
Washington, D. C. 20360

Dr. Neil McAlister
Office of Chief of Naval Operations
Command and Control
OP-094H
Washington, D. C. 20350

Naval Training Equipment Center
ATTN: Technical Library
Orlando, FL 32813

Department of the Navy

Dr. Robert G. Smith
Office of the Chief of Naval
Operations, OP987H
Personnel Logistics Plans
Washington, D. C. 20350

Combat Control Systems Department
Code 35
Naval Underwater Systems Center
Newport, RI 02840

Human Factors Department
Code N-71
Naval Training Equipment Center
Orlando, FL 32813

Dr. Alfred F. Smode
Training Analysis and Evaluation
Group
Naval Training & Equipment Center
Orlando, FL 32813

CDR Norman E. Lane
Code N-7A
Naval Training Equipment Center
Orlando, FL 32813

Dr. Gary Poock
Operations Research Department
Naval Postgraduate School
Monterey, CA 93940

Dean of Research Administration
Naval Postgraduate School
Monterey, CA 93940

Mr. H. Talkington
Engineering & Computer Science
Code 09
Naval Ocean Systems Center
San Diego, CA 92152

Human Factors Engineering
Code 8231
Naval Ocean Systems Center
San Diego, CA 92152

Department of the Navy

Mr. Paul Heckman
Naval Ocean Systems Center
San Diego, CA 92152

Dr. Ross Pepper
Naval Ocean Systems Center
Hawaii Laboratory
P. O. Box 997
Kailua, HI 96734

Dr. A. L. Slafkosky
Scientific Advisor
Commandant of the Marine Corps
Code RD-1
Washington, D. C. 20380

Dr. L. Chmura
Naval Research Laboratory
Code 7592
Computer Sciences & Systems
Washington, D. C. 20375

Office of the Chief of Naval
Operations (OP-115)
Washington, D.C. 20350

Professor Douglas E. Hunter
Defense Intelligence College
Washington, D.C. 20374

CDR C. Hutchins
Code 55
Naval Postgraduate School
Monterey, CA 93940

Human Factors Technology Administrator
Office of Naval Technology
Code MAT 0722
800 N. Quincy Street
Arlington, VA 22217

CDR Tom Jones
Naval Air Systems Command
Human Factors Programs
NAVAIR 330J
Washington, D. C. 20361

Department of the Navy

Commander
Naval Air Systems Command
Crew Station Design
NAVAIR 5313
Washington, D. C. 20361

Mr. Philip Andrews
Naval Sea Systems Command
NAVSEA 61R
Washington, D. C. 20362

Commander
Naval Electronics Systems Command
Human Factors Engineering Branch
Code 81323
Washington, D. C. 20360

Larry Olmstead
Naval Surface Weapons Center
NSWC/DL
Code N-32
Dahlgren, VA 22448

Mr. Milon Essoglou
Naval Facilities Engineering Command
R&D Plans and Programs
Code 03T
Hoffman Building II
Alexandria, VA 22332

CAPT Robert Biersner
Naval Medical R&D Command
Code 44
Naval Medical Center
Bethesda, MD 20014

Dr. Arthur Bachrach
Behavioral Sciences Department
Naval Medical Research Institute
Bethesda, MD 20014

Dr. George Moeller
Human Factors Engineering Branch
Submarine Medical Research Lab
Naval Submarine Base
Groton, CT 06340

Department of the Navy

Head
Aerospace Psychology Department
Code L5
Naval Aerospace Medical Research Lab
Pensacola, FL 32508

Commanding Officer
Naval Health Research Center
San Diego, CA 92152

Dr. Jerry Tobias
Auditory Research Branch
Submarine Medical Research Lab
Naval Submarine Base
Groton, CT 06340

Navy Personnel Research and
Development Center
Planning & Appraisal Division
San Diego, CA 92152

Dr. Robert Blanchard
Navy Personnel Research and
Development Center
Command and Support Systems
San Diego, CA 92152

CDR J. Funaro
Human Factors Engineering Division
Naval Air Development Center
Warminster, PA 18974

Mr. Stephen Merriman
Human Factors Engineering Division
Naval Air Development Center
Warminster, PA 18974

Mr. Jeffrey Grossman
Human Factors Branch
Code 3152
Naval Weapons Center
China Lake, CA 93555

Human Factors Engineering Branch
Code 4023
Pacific Missile Test Center
Point Mugu, CA 93042

Department of the Navy

Dean of the Academic Departments
U. S. Naval Academy
Annapolis, MD 21402

Dr. W. Moroney
Human Factors Section
Systems Engineering Test
Directorate
U. S. Naval Air Test Center
Patuxent River, MD 20670

Human Factor Engineering Branch
Naval Ship Research and Development
Center, Annapolis Division
Annapolis, MD 21402

Dr. Harry Crisp
Code N 51
Combat Systems Department
Naval Surface Weapons Center
Dahlgren, VA 22448

Mr. John Quirk
Naval Coastal Systems Laboratory
Code 712
Panama City, FL 32401

Department of the Army

Dr. Edgar M. Johnson
Technical Director
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Technical Director
U.S. Army Human Engineering Labs
Aberdeen Proving Ground, MD 21005

Director, Organizations and
Systems Research Laboratory
U. S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Department of the Army

Mr. J. Barber
HQS, Department of the Army
DAPE-MBR
Washington, D. C. 20310

Department of the Air Force

U. S. Air Force Office of Scientific Research
Life Sciences Directorate, NL
Bolling Air Force Base
Washington, D. C. 20332

AFHRL/LRS TDC
Attn: Susan Ewing
Wright-Patterson AFB, OH 45433

Chief, Systems Engineering Branch
Human Engineering Division
USAF AMRL/HES
Wright-Patterson AFG, OH 45433

Dr. Earl Alluisi
Chief Scientist
AFHRL/CCN
Brooks Air Force Base, TX 78235

Dr. R. K. Dismukes
Associate Director for Life Sciences
AFOSR
Bolling AFB
Washington, D.C. 20332

Foreign Addresses

Dr. Kenneth Gardner
Applied Psychology Unit
Admiralty Marine Tech. Estab.
Teddington, Middlesex TW11 OLN
England

Dr. Daniel Kahneman
University of British Columbia
Department of Psychology
Vancouver, BC V6T 1W5
Canada

Foreign Addressees

Human Factors
P. O. Box 1085
Station B
Rexdale, Ontario
Canada M9V 2B3

Dr. A. D. Baddeley
Director, Applied Psychology Unit
Medical Research Council
15 Chaucer Road
Cambridge, CB2 2EF England

Other Government Agencies

Defense Technical Information Center
Cameron Station, Bldg. 5
Alexandria, VA 22314 (12 copies)

Dr. Clinton Kelly
Defense Advanced Research Projects Agency
1400 Wilson Blvd.
Arlington, VA 22209

Dr. M. D. Montemerlo
Human Factors & Simulation Technology, RTE-6
NASA HQS
Washington, D.C. 20546

Other Organizations

Ms. Denise Benel
Essex Corporation
333 N. Fairfax Street
Alexandria, VA 22314

Dr. Andrew P. Sage
School of Engineering and Applied Science
University of Virginia
Charlottesville, VA 22901

Other Organizations

Dr. Robert R. Mackie
Human Factors Research Division
Canyon Research Group
5775 Dawson Avenue
Goleta, CA 93017

Dr. Amos Tversky
Department of Psychology
Stanford University
Stanford, CA 94305

Dr. H. McI. Parsons
Essex Corporation
333 N. Fairfax
Alexandria, VA 22314

Dr. Jesse Orlansky
Institute for Defense Analyses
1801 N. Beauregard Street
Alexandria, VA 22311

Dr. J. O. Chinnis, Jr.
Decision Science Consortium, Inc.
7700 Leesburg Pike
Suite 421
Falls Church, VA 22043

Dr. T. B. Sheridan
Department of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge, MA 02139

Dr. Paul E. Lehner
PAR Technology Corp.
P.O. Box 2005
Reston, VA 22090

Dr. Paul Slovic
Decision Research
1201 Oak Street
Eugene, OR 97401

Other Organizations

Dr. Harry Snyder
Dept of Industrial Engeering
Virginia Polytechnic Institute
and State University
Blacksburg, VA 24061

Dr. Stanley Deutsch
NAS-National Research Council (COHF)
2101 Constitution Avenue, N.W.
Washington, D.C. 20418

Dr. Amos Freedy
Perceptronics, Inc.
6271 Variel Avenue
Woodland Hills, CA 91364

Dr. Robert Fox
Department of Psychology
Vanderbilt University
Nashville, TN 37240

Dr. Meredith P. Crawford
American Psychological Association
Office of Educational Affairs
1200 17th Street, N.W.
Washington, D.C. 20036

Dr. Deborah Boehm-Davis
General Electric Company
Information & Data Systems
1755 Jefferson Davis Highway
Arlington, VA 22202

Dr. Howard E. Clark
NAS-NRC
Commission on Engrg. & Tech. Systems
2101 Constitution Ave., N.W.
Washington, D.C. 20418

Other Organizations

Dr. Charles Gettys
Department of Psychology
University of Oklahoma
455 West Lindsey
Norman, OK 73069

Dr. Kenneth Hammond
Institute of Behavioral Science
University of Colorado
Boulder, CO 80309

Dr. James H. Howard, Jr.
Department of Psychology
Catholic University
Washington, D. C. 20064

Dr. William Howell
Department of Psychology
Rice University
Houston, TX 77001

Dr. Christopher Wickens
Department of Psychology
University of Illinois
Urbana, IL 61801

Mr. Edward M. Connelly
Performance Measurement
Associates, Inc.
410 Pine Street, S. E.
Suite 300
Vienna, VA 22180

Professor Michael Athans
Room 35-406
Massachusetts Institute of
Technology
Cambridge, MA 02139

Dr. Edward R. Jones
Chief, Human Factors Engineering
McDonnell-Douglas Astronautics Co.
St. Louis Division
Box 516
St. Louis, MO 63166

Other Organizations

Dr. Babur M. Pulat
Department of Industrial Engineering
North Carolina A&T State University
Greensboro, NC 27411

Dr. Lola Lopes
Information Sciences Division
Department of Psychology
University of Wisconsin
Madison, WI 53706

National Security Agency
ATTN: N-32, Marie Goldberg
9800 Savage Road
Ft. Meade, MD 20722

Dr. Stanley N. Roscoe
New Mexico State University
Box 5095
Las Cruces, NM 88003

Mr. Joseph G. Wohl
Alphatech, Inc.
3 New England Executive Park
Burlington, MA 01803

Dr. Marvin Cohen
Decision Science Consortium, Inc.
Suite 721
7700 Leesburg Pike
Falls Church, VA 22043

Dr. Robert Wherry
Analytics, Inc.
2500 Maryland Road
Willow Grove, PA 19090

Dr. William R. Uttal
Institute for Social Research
University of Michigan
Ann Arbor, MI 48109

Dr. William B. Rouse
School of Industrial and Systems
Engineering
Georgia Institute of Technology
Atlanta, GA 30332

Other Organizations

Dr. Richard Pew
Bolt Beranek & Newman, Inc.
50 Moulton Street
Cambridge, MA 02238

Dr. Hillel Einhorn
Graduate School of Business
University of Chicago
1101 E. 58th Street
Chicago, IL 60637

Dr. Douglas Towne
University of Southern California
Behavioral Technology Lab
3716 S. Hope Street
Los Angeles, CA 90007

Dr. David J. Getty
Bolt Beranek & Newman, Inc.
50 Moulton street
Cambridge, MA 02238

Dr. John Payne
Graduate School of Business
Administration
Duke University
Durham, NC 27706

Dr. Baruch Fischhoff
Decision Research
1201 Oak Street
Eugene, OR 97401

Dr. Alan Morse
Intelligent Software Systems Inc.
160 Old Farm Road
Amherst, MA 01002

Dr. J. Miller
Florida Institute of Oceanography
University of South Florida
St. Petersburg, FL 33701

END

FILMED

8

DTIC